Differential evolution classifier with optimized distance measures from a pool of distances

Author(s)

David Koloseni ; Jouni Lampinen ; Pasi Luukka

Abstract:

In this article we propose a differential evolution based nearest prototype classifier with extension to selecting the applied distance measure from a pool of alternative measures optimally for the particular data set at hand. The proposed method extends the earlier differential evolution based nearest prototype classifier by extending the optimization process to cover also the selection of distance measure instead of optimizing only the parameters related with a preselected and fixed distance measure. Now the optimization process is seeking also for the best distance measure providing the highest classification accuracy over the selected data set. It has been clear for some time that in classification, the usual euclidean distance measure is sometimes not the best possible choice. Still usually not much has been done for it, and in many cases where some consideration to this problem is given, there has only been testing with a couple of alternative distance measures to find which one provides the highest classification accuracy over the current data set. In this paper we attempt to take one step further by not only enumerating a couple of alternative distance measures, but applying a systematic optimization process to select the best distance measure from a pool of multiple alternative distance measures. In parallel, within the same optimization process, the optimal parameter values related to each alternative distance measures are determined as well as the optimal class prototype vectors for the given data. The empirical results represented are indicating that with several data sets the optimal distance measure is some other measure than the most commonly applied euclidean distance. The results are also suggesting that from the classification accuracy point of view the proposed global optimization approach has high potential in solving classification problems of the studied type. Perhaps the most generally applicable conclusion from our results is, that emphasizing of selection of distance measure is more important to classification accuracy that it has been commonly believed so far.